LAW OFFICES
McGuireWoods LLP
1750 TYSONS BOULEVARD, SUITE 1800
MCLEAN, VIRGINIA 22102

APPLICATION
FOR
UNITED STATES
LETTERS PATENT

Applicants: Se June Hong, Ramesh Natarajan and
        Ilana Belitskaya
For: FAST METHOD FOR RENEWAL AND
        ASSOCIATED RECOMMENDATIONS FOR
        MARKET BASKET ITEMS
Docket No.: YOR9-2000-0776US

FAST METHOD FOR RENEWAL AND ASSOCIATED RECOMMENDATIONS FOR MARKET BASKET ITEMS

DESCRIPTION

BACKGROUND OF THE INVENTION

5

Field of the Invention

The present invention generally relates to a computer method and system for placing orders for products over a computer network, such as the Internet, and more particularly, to a way to more effectively and efficiently determine a customer's preferences while the customer's choices are in

10 progress in order to make recommendations of other items the customer might be interested in purchasing. More generally, further recommendations while a customer is making choices applies to any such situation, e.g., a customer makes a series of Internet surfing choices and new sites are dynamically recommended and displayed (by icons). Aside from virtual shopping carts,

15 this can also apply to the real shopping cart with displays. As a customer fills the cart, the display points to the next items the customer is likely to add to the cart.

Background Description

Shopping on the World Wide Web (WWW or simply the Web) portion

20 of the Internet has become ubiquitous in our society. A typical Web site offering products for purchase employs what is referred to as a "market basket", a sort of virtual shopping cart without wheels. The customer selects

items to add to his or her market basket, and when he or she completes their shopping, a "check out" button is selected to process the items then in the market basket.

5   A market strategy has developed which involves monitoring the items in the customer's market basket and, taking other factors into account including possibly the customer's past buying habits and similar choices made by other customers, making recommendations to the customer of other items he or she might be interested in purchasing. In the past decade, recommendations to a customer who has items in a market basket have been

10   made using so called associative rules mined from the market basket data, or by several other means described, for example, in P. Resnick, N. Iacovou, M. Suchak, P. Berstrom and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews", *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work*, pp. 175–186, ACM, New York

15   (1994), J. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", *Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufman, Madison, Wisc. (1998), and others. The associative rules cannot be tailored to all possible partial market baskets. All the prior art in so called collaborative

20   filtering technique require a substantial amount of computation.

SUMMARY OF THE INVENTION

        It is therefore an object of the present invention to provide a more effective and efficient process for recommending items to a customer for their market basket in an e-commerce site.

25   According to the invention, a new method is provided which is based on a novel theory that "not all items in the basket are selected because of their affinity with some other item already in the basket." The method uniquely

determines two separate components of item choice preferences: Preference by association with existing items in the basket in progress or independently exercised purchases. The former is the usual preference considered by all prior art methods. The latter is the renewal buying not considered by the prior art. In the present invention, these two preferences are separately estimated from the training data and combined in proper proportions to obtain the overall preference for each item not yet in the basket. The recommendations are presented in the form of ranking from which some subset of items at the top will be presented to the customer. The ranking is obtained from computed probabilities for each item that is not in the current basket, given the partial basket in progress. The method disclosed here is not restricted to purchasing of items. It can also be used for recommending new web-sites to someone browsing the Internet.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, aspects and advantages will be better understood from the following detailed description of a preferred embodiment of the invention with reference to the drawings, in which:

Figure 1 is a table showing an array of binary data which represents items in market baskets; and

Figure 2 is a flow diagram showing the logic of the computer implemented process according to the invention.

·

DETAILED DESCRIPTION OF A PREFERRED
EMBODIMENT OF THE INVENTION

Referring now to the drawings, and more particularly to Figure 1, there is shown a table which illustrates a binary array which represents items in

market baskets. In this table, each row is a basket and each column represents an item. A binary value 1 in row i and column j signifies that the basket i contained item j and value 0 for the absence of the item. We shall denote such market basket data array as $\mathbf{M}$, comprised of n baskets (rows) and m items

5   (columns).

The current partial basket is denoted as $\mathbf{B}$, the content items of which is denoted as $i_1, i_2, \ldots, i_b$, where the number of items in the basket, b, can be 0 if the basket is just beginning. Such case will be called a null basket, and the method to determine the preferences for the null basket will be separately

10   described later.

The probability of a customer buying item j given the partial basket $\mathbf{B}$ is $P(j|\mathbf{B})$. The key concept is to separately consider the probability components: one due to associative buying, and the other due to an independent, or renewal choice.

15     $$\mathbf{P}(j|\mathbf{B}) = \mathbf{P}(j, \text{asso} \mid \mathbf{B}) + \mathbf{P}(j, \text{renewal} \mid \mathbf{B})$$
$$= \mathbf{P}(j| \text{asso}, \mathbf{B}) \, \mathbf{P}(\text{asso} \mid \mathbf{B}) + \mathbf{P}(j \mid \text{renewal}, \mathbf{B}) \, \mathbf{P}(\text{renewal} \mid \mathbf{B}), \qquad (1)$$

for all j not in $\mathbf{B}$ where, since one buys associatively or independently,

$$\mathbf{P}(\text{asso} \mid \mathbf{B}) = 1 - \mathbf{P}(\text{renewal} \mid \mathbf{B}) \qquad (2)$$

And in the case of renewal buy, the basket content is immaterial except for

20   those items already in the partial basket $\mathbf{B}$, and hence

$$\mathbf{P}(j \mid \text{renewal}, \mathbf{B}) = \mathbf{P}(j \mid \text{renewal}) = \mathbf{P}(j, \text{renewal})/\mathbf{P}(\text{renewal})$$
$$= \mathbf{P}(\text{renewal} \mid j) \, \mathbf{P}(j)/\mathbf{P}(\text{renewal}), \qquad (3)$$

where $\mathbf{P}(j)$ is the probability of item j being bought.

YOR9-2000-0776US

Now we make a simple but reasonable assumption about the purchase behavior we name "single item influence". That is, whether the next buy is renewal or associative, it is determined as an aggregate of such tendency by the items in the current basket, singly. In other words, an associative next buy

5 would be the result of its association to some one item in the basket and not because more than one item was needed for the association. We, likewise, assume each single item exerts its own tendency to non-associative, i.e., renewal, buying. These assumptions are reasonable and allow an efficient computation.

10 We make further simplifying assumptions about the purchasing behavior regarding the aggregation of the single item influence. In the case of renewal, we reasonably assume that the least renewal tendency among all the basket items dictate the final renewal. So, for aggregating the renewal probabilities,

15 $$\mathbf{P}(\text{renewal} \mid \mathbf{B}) = \min_k \mathbf{P}(\text{renewal} \mid i_k), \text{ for } k = 1, 2, \ldots, b, \tag{4}$$

which will be estimated from the data in a manner described below. And in the case of associated buying, we reasonably assume that maximum preference to associatively select an item j among each item in the partial basket $\mathbf{B}$ determines the overall preference for the item j. That is, in pre-

20 normalized form,

$$\mathbf{P}'(j \mid \text{asso}, \mathbf{B}) = \max_k \mathbf{P}(j \mid \text{asso}, i_k) \text{ for } k = 1, 2, \ldots, b. \tag{5}$$

This quantity is set to zero for all items in the future partial basket for which recommendations are made. After that, they are normalized for probability, as

$$\mathbf{P}(j \mid \text{asso},\mathbf{B}) = \frac{\mathbf{P}'(j \mid \text{asso},\mathbf{B})}{\sum_{j} \mathbf{P}'(j \mid \text{asso},\mathbf{B})} \quad \text{for all items } j. \tag{6}$$

Now, the probability, $\mathbf{P}(j \mid \text{asso}, i_k)$, of equation (5) is equivalent to (using i for i_k)

$$\mathbf{P}(j \mid \text{asso},i) = \frac{\mathbf{P}(j,i,\text{asso})}{\mathbf{P}(i,\text{asso})} = \frac{\mathbf{P}(j,i)\mathbf{P}(\text{asso} \mid j,i)}{\mathbf{P}(i)\mathbf{P}(\text{asso} \mid i)}$$

$$= \left\{\frac{\mathbf{P}(j,i)}{\mathbf{P}(i)}\right\} \frac{\{1 - \mathbf{P}(\text{renewal} \mid j,i)\}}{\{1 - \mathbf{P}(\text{renewal} \mid i)\}}$$

$$= \frac{\mathbf{P}(j \mid i)\,\{1 - \mathbf{P}(\text{renewal} \mid j,i)\}}{\{1 - \mathbf{P}(\text{renewal} \mid i)\}} \tag{7}$$

When the partial basket in progress is empty, i.e., the null basket at the start, a customer is at precisely the "renewal" point. Therefore, for null basket $\mathbf{B} = \text{null}$, equation (1) is specialized by use of equation (3).

$$\mathbf{P}(j \mid \text{null}) = \mathbf{P}(j \mid \text{renewal}) \tag{8}$$

Now we describe sub methods to estimate $\mathbf{P}(j)$, $\mathbf{P}(\text{renewal})$, $\mathbf{P}(\text{renewal} \mid j,i\}$, and $\mathbf{P}(j \mid i)$, etc. of the above equations from the data.

- $\mathbf{P}(j)$ estimation: precomputed and stored in length m vector.

Let the column sums of $\mathbf{M}$ be $n_1, n_2, \ldots, n_k, \ldots, n_m$. The probability of item j being bought is then as

YOR9-2000-0776US

$$P(j) = \frac{n_j}{\sum\limits_k n_k} \qquad (9A)$$

or optionally with a Laplace correction for small statistics as

$$P(j) = \frac{n_j + 1}{\sum\limits_k n_k + 1} \qquad (9B)$$

- P(renewal) estimation.

Let the number of singleton baskets of item j be n_j'. This quantity is underestimated by the proportion of all singleton baskets to the total items purchased in the training data. The reason is that every time only one item was bought, it is certainly a case of renewal. The renewal probability is then

$$P(\text{renewal}) = \frac{\sum\limits_j n_j'}{\sum\limits_j n_j} \qquad (10)$$

- P(renewal | i) estimation: precomputed and stored in a length m vector.

Given the item i is bought, the estimate of renewal probability is done in two stages. Let the total number of baskets where the item i is the singleton basket content be n_i', then for $\dfrac{n_i'}{n_i}$ of the time, it is certain case of renewal,

and for the remaining proportions, i.e., for $1-\left(\dfrac{n_i{}'}{n_i}\right)$ of the time, there are

other items bought along with the item i, but some portion of it, which we

estimate to be **P**(renewal), would be also renewal case. Therefore, the estimate

is

5
$$P\,(\text{renewal} \mid i) \;=\; \left(\frac{n_i{}'}{n_i}\right) + \mathbf{P}(\text{renewal}) \times \left(1 - \left(\frac{n_i{}'}{n_i}\right)\right) \tag{11}$$

- **P**(j | renewal) computation: precomputed and stored in a length m

  vector.

  P(j | renewal) is computed using the above estimated quantities and

  stored according to equation (3).

10
- **P**(j | i) estimation:

  Let the subset of **M** that has 1 in i-th column be $\mathbf{M}_i$, i.e., those rows

  that have item i in the basket. The j-th column sum of $\mathbf{M}_i$, denoted as n_{ji},

  represent the number of times j was bought along with i. Therefore,

$$\mathbf{P}(j \mid i) \;=\; \frac{n_{ji}}{\displaystyle\sum_{k} n_{ki}}\;, \text{ and we fix } \mathbf{P}(i \mid i) \text{ to be } 0 \tag{12}$$

15
- **P**(renewal | j,i) estimation:

  From sub matrix $\mathbf{M}_i$ above, the number of rows whose sum is exactly 2

  represents a certain case of renewal. Let $n_{ji}{}'$ denote the number of rows whose

row sum in $\mathbf{M}_i$ is exactly 2 and contains item j. The certain renewal porportion is n_{ji}' / n_{ji}. In the remaining cases, we estimate that the renewal is the same as $\mathbf{P}$(renewal). So,

$$P(\text{renewal} \mid j,i) = \left(\frac{n_{ji}'}{n_{ji}} \right) + P(\text{renewal}) \times \left(1 - \left(\frac{n_{ji}'}{n_{ji}} \right) \right) \tag{13}$$

5 • $\mathbf{P}(j \mid \text{asso}, i)$ computation: precomputed and stored in an m by m array or an equivalent sparce matrix representation.

Using the estimate above, $\mathbf{P}(j \mid \text{asso}, i)$ of equation (7) is computed and stored.

• $\mathbf{P}(j \mid \text{asso}, \mathbf{B})$ computation:

10 First, we obtain $\mathbf{P}'(j \mid \text{asso}, \mathbf{B})$ of equation (5) using equation (7) and the quantities developed above. Since the items already in the partial basket are not bought again, we fix it to zero whenever j is in $\mathbf{B}$. Now, the normalized probability of j being purchased associated with the partial basket is

$$P(j \mid \text{asso}, \mathbf{B}) = \frac{\mathbf{P}'(j \mid \text{asso}, \mathbf{B})}{\sum_k \mathbf{P}'(k \mid \text{asso}, \mathbf{B})} \tag{14}$$

15 • $\mathbf{P}(j \mid \text{renewal}, \mathbf{B}) = \mathbf{P}(j \mid \text{renewal})$ normalization for partial basket $\mathbf{B}$:

The $\mathbf{P}(j \mid \text{renewal}, \mathbf{B}) = \mathbf{P}(j \mid \text{renewal})$ of equation (3) is now fixed for those j's that are already in the partial basket B to be zero, and normalized by dividing them by the sum over all j's before the final goal $\mathbf{P}(j \mid \mathbf{B})$ is computed

from equation (1) using the partial quantities developed herewith.

The final recommendation for items based on the current partial basket in progress is then in descending $P(j \mid B)$ ranking. The probability itself can be used for a direct gain maximization if the profit amount for each item is known. It that case, one would multiply the probabilities with the corresponding profit amount before ranking is made. More specifically, when each item's profit amount, $\$_j$, is known, one computes $P(j \mid B)\$_j$ and produces the ranking for recommendations based on this quantity.

The process is illustrated in Figure 2. The method comprises three steps. The first two steps use the market basket information in the training data base 201. Specifically, in the first step 202, certain statistics are collected which are then used in the second step 203 to precompute certain quantities. The third step 204 uses the precomputed quantities, in the stored statistical model 205, and the partial market basket information 206 in an online manner to produce a preference ranking for the remaining unpurchased items. We assume the training data to contain n market baskets with m items.

In more detail, the first step 202 is to collect statistics from the training data. This involves the following:

(a)   For each item j, obtain n_j the number of baskets with item j purchased.

(b)   For each item j, obtain n_j' the number of baskets with j being the sole item purchased.

(c)   For each pair of items i and j, obtain the number of market baskets n_{ji} with items j and i purchased together.

(d)   For each pair of items i and j, obtain the number of market baskets n_{ji}' with items i and j being the only two items purchased.

The second step 203 is to precompute model parameters. This involves the following:

(a) Compute $P(renewal) = \dfrac{\sum_k n_k'}{\sum_k n_k}$ (equation (10)).

(b) For each item j, compute $P(j) = \dfrac{n_j}{\sum_k n_k}$ (equation (9A), or use

equation (9B)).

(c) For each item j, compute $P(renewal \mid j) = \dfrac{n_j'}{n_j} + P(renewal)\left(1 - \dfrac{n_j'}{n_j}\right)$

5   (equation (11)).

(d) For each item j, compute

$P'(j \mid renewal) = P(renewal \mid j) \times \dfrac{P(j)}{P(renewal)}$ (equation (3)).

(e) For each pair of items i and j with $n_{ij} \neq 0$, compute $P(j \mid i) = \dfrac{n_{ji}}{\sum_k n_{ki}}$

(equation (12)).

10   (f) For each pair of items i and j with $n_{ij} \neq 0$, compute

$P(renewal \mid j,i) = \dfrac{n_{ji}'}{n_{ji}} + P(renewal)\left(1 - \dfrac{n_{ji}'}{n_{ji}}\right)$ (equation (13)).

(g) For each pair of items i and j with $n_{ij} \neq 0$, compute

$P'(j \mid asso,i) = P(j \mid i) \times \dfrac{(1 - P(renewal \mid j,i))}{(1 - P(renewal \mid i))}$ (equation (7)).

The third step is to calculate a recommended ordering for a given partial

YOR9-2000-0776US

market basket. Given a partial basket $\mathbf{B} = \{i_1, i_2, \ldots, i_k\}$, let $\overline{\mathbf{B}}$ be the complementary set of items not in $\mathbf{B}$. Then

(a) If $\mathbf{B}$ is empty, the sort items in order of decreasing $\mathbf{P}(j \mid \text{renewal})$ and return this as the item preference ordering.

(b) If $\mathbf{B}$ is non-empty, then

(i) Compute $\mathbf{P}(\text{renewal} \mid \mathbf{B}) = \min_{i_k \in \mathbf{B}} \mathbf{P}(\text{renewal} \mid i_k)$ (equation (4)).

(ii) Compute the normalization factor $\displaystyle\sum_{k \in \overline{\mathbf{B}}} \mathbf{P}'(k \mid \text{renewal})$.

(iii) For each item $j \in \overline{\mathbf{B}}$, compute

$$\mathbf{P}(j \mid \text{renewal}) = \frac{\mathbf{P}'(j \mid \text{renewal})}{\displaystyle\sum_{k \in \overline{\mathbf{B}}} \mathbf{P}'(k \mid \text{renewal})} \ .$$

(iv) Compute the normalization factor $\displaystyle\sum_{j \in \overline{\mathbf{B}}} \mathbf{P}'(j \mid \text{asso,}\mathbf{B})$.

(v) For each item $j \in \overline{\mathbf{B}}$, compute

$$\mathbf{P}'(j \mid \text{asso,}\mathbf{B}) = \max_{i_k \in \mathbf{B}} \mathbf{P}(j \mid \text{asso,}i_k) \text{ (equation (5))}.$$

(vi) For each item $j \in \overline{\mathbf{B}}$, compute

$$\mathbf{P}(j \mid \text{asso,}\mathbf{B}) = \frac{\mathbf{P}'(j \mid \text{asso,}\mathbf{B})}{\displaystyle\sum_{k \in \overline{\mathbf{B}}} \mathbf{P}'(k \mid \text{asso,}\mathbf{B})} \text{ (equation (6))}.$$

(vii) For each item $j \in \overline{\mathbf{B}}$, compute

$\mathbf{P}(j|\mathbf{B}) = \mathbf{P}(j \mid \text{asso,}\mathbf{B})\mathbf{P}(\text{asso} \mid \mathbf{B}) + \mathbf{P}(j \mid \text{renewal,}\mathbf{B})\mathbf{P}(\text{renewal} \mid \mathbf{B})$ (equation (1)).

(viii) Sort items in order of decreasing $\mathbf{P}(j \mid \mathbf{B})$ and return this as the

item preference ordering.

One skilled in the art can utilize many techniques to reduce the storage requirement to process the present invention when the number of items is very large: reduced accuracy for probabilities, sparce matrix storing techniques, and clustering of like items to reduce the number of items, which can be later refined for the cluster members after the cluster preferences are computed.

While the invention has been described in terms of a single preferred embodiment, those skilled in the art will recognize that the invention can be practiced with modification within the spirit and scope of the appended claims.

YOR9-2000-0776US